

Estimation of Spectral Risk Measures

Ajay Kumar Pandey,¹ Prashanth L. A.,¹ Sanjay P. Bhat²

¹ Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, India

² TCS Research, Hyderabad, India, and Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, India

ajaykp@cse.iitm.ac.in, prashla@cse.iitm.ac.in, sanjay.bhat@tcs.com

Abstract

We consider the problem of estimating a spectral risk measure (SRM) from i.i.d. samples, and propose a novel method that is based on numerical integration. We show that our SRM estimate concentrates exponentially when the underlying distribution has bounded support. Further, we also consider the case when the underlying distribution satisfies an exponential moment bound, which includes sub-Gaussian and sub-exponential distributions. For these distributions, we derive a concentration bound for our estimation scheme. We validate the theoretical findings on a synthetic setup, and in a vehicular traffic routing application.

1 Introduction

Conditional value-at-risk (CVaR) is a popular measure in the context of risk-sensitive optimization. CVaR is the expectation of a random variable (r.v.) (usually the loss in a financial application) conditioned on the r.v. exceeding the value-at-risk (VaR). The latter is a high-probability upper bound on the random loss that could be incurred. The advantage of employing CVaR instead of VaR in a risk-sensitive optimization setting is that CVaR is a coherent risk measure (Artzner et al. 1999), while VaR is not.

Spectral risk measures (SRM), proposed in (Acerbi 2002), generalize VaR and CVaR. SRM uses a ‘risk-aversion’ function to assign a weight to each loss. Note that VaR gives a zero weight for each loss beyond a given quantile, while CVaR assigns a constant weight in the tail region beyond VaR. In contrast, SRMs can assign larger weights to higher losses, and thus model a user’s risk attitude better. Moreover, SRMs satisfy coherency provided the risk-aversion function is positive, increasing and integrates to one.

In this paper, we consider the problem of estimating the SRM of a r.v., given independent and identically distributed (i.i.d.) samples from the underlying distribution. In this context, our contributions are as follows: First, we provide a natural estimation scheme for SRM that uses the empirical distribution function (EDF) to estimate quantiles, together with a trapezoidal rule-based approximation to the integral in the definition of the SRM. Second, we provide concentration bounds for our proposed SRM estimate for the following two

cases: first, when the underlying distribution is assumed to have bounded support; and second, when the distribution satisfies an exponential moment bound. The latter class includes sub-Gaussian and sub-exponential distributions (Wainwright 2019). Third, we perform simulation experiments to show the efficacy of our proposed SRM estimation scheme. In particular, we consider a synthetic setup, and show that our scheme provides accurate estimates of SRM. Next, we incorporate our SRM estimation scheme in the inner loop of the successive rejects (SR) algorithm (Audibert, Bubeck, and Munos 2010), which is a popular algorithm in the best arm identification framework for multi-armed bandits. We test the resulting SR algorithm variant in a vehicular traffic routing application using SUMO traffic simulator (Behrisch et al. 2011). The application is motivated by the fact that, in practice, human road users may not always prefer the route with lowest mean delay. Instead, a route that minimizes worst-case delay, while doing reasonably well on the average, is preferable, and such a preference can be encoded into the risk aversion function employed in SRMs.

Related work. In (Brown 2007; Wang and Gao 2010) concentration bounds for the classic CVaR estimator are derived assuming that the underlying distribution has bounded support. Our bound, using a different estimator, exhibits a similar rate of exponential convergence around true CVaR. For the case of distributions with unbounded support, concentration bounds for empirical CVaR have been derived recently in (Kolla et al. 2019; Kagracha, Nair, and Jagannathan 2019; Bhat and Prashanth 2019; Prashanth, Jagannathan, and Kolla 2020). In (Kolla et al. 2019) (resp. (Bhat and Prashanth 2019; Prashanth, Jagannathan, and Kolla 2020)), the authors derive a one-sided concentration bound (resp. two-sided bounds), when the underlying distributions are either sub-Gaussian or sub-exponential. In comparison to (Kolla et al. 2019), we derive two-sided concentration results, and our bounds show an improved dependence on the number of samples, say n , and accuracy, say ϵ . In (Khim et al. 2020), the authors employ tools from statistical learning theory to derive a concentration result for a class of risk measures that includes CVaR under the assumption that the underlying loss distribution has bounded support.

The closest related work is (Bhat and Prashanth 2019), where the authors employ a Wasserstein distance approach to provide concentration bounds for SRM estimation. In com-

parison to (Bhat and Prashanth 2019), our bounds have a similar dependence on n and ϵ for the sub-Gaussian case, and a much improved dependence on ϵ for the sub-exponential case. Further, unlike (Bhat and Prashanth 2019), we specify all the constants in our bounds.

From an application viewpoint, CVaR-based models have been explored in different contexts, for instance, in a bandit application (Galichet, Sebag, and Teytaud 2013), in a portfolio optimization problem (Krokhmal, Palmquist, and Uryasev 2002), and in a general risk management setting (Mulvey and Erkan 2006). In the simulation experiments, we consider a bandit application in the context of vehicular traffic routing, with SRM as the performance objective, and show the efficacy of our proposed estimation scheme.

2 Preliminaries

For a r.v. X , VaR $V_\beta(X)$ and CVaR $C_\beta(X)$ at the level β , $\beta \in (0, 1)$, are defined as follows:

$$V_\beta(X) := \inf\{c : P(X \leq c) \geq \beta\}, \quad (1)$$

$$C_\beta(X) := V_\beta(X) + \frac{1}{1-\beta} \mathbb{E}[X - V_\beta(X)]^+, \quad (2)$$

where $[x]^+ = \max(0, x)$ for a real number x . $V_\beta(X)$ can be interpreted as the smallest loss that will not be exceeded with probability β . Note that, if X has a continuous and strictly increasing cumulative distribution function (CDF) F , then $V_\beta(X) = F^{-1}(\beta)$. Further, $C_\beta(X)$ can be interpreted as the expected loss, conditional on the event that the loss exceeds $V_\beta(X)$, i.e., $C_\beta(X) = \mathbb{E}[X | X \geq V_\beta(X)]$. Acerbi's formula (Acerbi and Tasche 2002), an alternative form for $C_\alpha(X)$ defined in (2), is as follows:

$$C_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 V_\beta(X) d\beta. \quad (3)$$

Let $X_i, i = 1, \dots, n$ denote i.i.d. samples from the distribution of X . Then, the estimate of $V_\beta(X)$, denoted by $\widehat{V}_{n,\beta}$, is formed as follows (Serfling 2009):

$$\widehat{V}_{n,\beta} = \widehat{F}_n^{-1}(\beta) = \inf\{x : \widehat{F}_n(x) \geq \beta\}, \quad (4)$$

where $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[X_i \leq x]$ is the EDF of X . Here $\mathbb{I}\{\cdot\}$ denotes the indicator function, i.e., for an event A , $\mathbb{I}\{A\} = 1$ if A happens, and $\mathbb{I}\{A\} = 0$ otherwise. Letting $X_{(1)}, \dots, X_{(n)}$ denote the order statistics, i.e., $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, we have $\widehat{V}_{n,\beta} = X_{(\lceil n\beta \rceil)}$.

3 Spectral Risk Measure (SRM)

Let $\mathcal{L}^1([0, 1])$ denote the space of equivalence classes of Lebesgue-integrable functions on $[0, 1]$, where two functions are considered equivalent if they agree everywhere except possibly on a set of zero Lebesgue measure. For any $f \in \mathcal{L}^1([0, 1])$, we define the following norm: $\|f\| = \int_0^1 |f(p)| dp$.

The definition of SRM requires a risk-aversion function $\varphi \in \mathcal{L}^1([0, 1])$, which is said to be *admissible* if it satisfies the following properties (see Definition 2.4 in (Acerbi 2002)): (i) φ is positive; (ii) φ is increasing; and (iii) $\|\varphi\| = 1$.

The SRM of a r.v. X is defined by

$$S(X) = \int_0^1 \varphi(\beta) V_\beta(X) d\beta, \quad (5)$$

where V_β is as defined in (1), and $\varphi \in \mathcal{L}^1([0, 1])$ is admissible. In Theorem 4.1 of (Acerbi 2002), the author establishes that admissibility of φ is both necessary and sufficient for ensuring coherency of $S(X)$.

Note that by virtue of working with the Lebesgue space $\mathcal{L}^1([0, 1])$, our conclusions remain unaffected if V_β is arbitrarily defined at $\beta = 0$ and $\beta = 1$. On the other hand, using the Riemann integral in (5) would have required us to restrict our attention to the open interval $(0, 1)$ to get around the fact that $\lim_{\beta \rightarrow 0} V_\beta = -\infty$.

We choose to focus on the family of r.v.s for which $S(X)$ is well-defined, i.e., $\{X \mid \varphi(\cdot) V_{(\cdot)}(X) \in \mathcal{L}^1([0, 1])\}$. As noted in (Acerbi 2002), if φ is bounded, and X is integrable, then $S(X)$ is well-defined.

SRM can be seen as a weighted average of the quantiles (VaR) of the underlying distribution, with the risk-aversion function φ used to define the weights. Moreover, CVaR can be recovered by using the following risk-aversion function:

$$\varphi(\beta) = \frac{1}{1-\alpha} \mathbb{I}\{\beta > \alpha\}, \quad \alpha \in (0, 1). \quad (6)$$

In particular, using (6) in (5) leads to (3). Notice that CVaR uses the same weight for all tail-loss VaR values. In contrast, the risk-aversion function φ can be chosen such that higher losses receive a higher weight or at least, the same weight as lower losses (Dowd and Blake 2006). Thus, SRM can model a user's risk aversion better.

4 SRM Estimation: Bounded Case

Estimation Scheme

We estimate $S(X)$, given i.i.d. samples X_1, \dots, X_n from the distribution of X , by approximating the integral in SRM definition (5). Notice that the integrand $V_\beta(X)$ in (5) has to be estimated using the samples. Recall that $\widehat{V}_{n,\beta}$ is the estimate of $V_\beta(X)$, given by (4). We use the weighted VaR estimate to form a discrete sum to approximate the integral, an idea motivated by the trapezoidal rule (Talvila 2016). The estimate $\widehat{S}_{n,m}$ of $S(X)$ is formed as follows:

$$\widehat{S}_{n,m} = \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) \widehat{V}_{n,\beta_{k-1}} + \varphi(\beta_k) \widehat{V}_{n,\beta_k}}{2} \Delta\beta. \quad (7)$$

In the equation above, $\{\beta_k\}_{k=0}^m$ is a partition of $[0, 1]$ such that $\beta_0 = 0$ and $\beta_k = \beta_{k-1} + \Delta\beta$, where $\Delta\beta = 1/m$ is the length of each sub-interval.

For the special case of $\varphi(\cdot)$ as defined in (6), we obtain an estimate of CVaR from (7) as follows:

$$\widehat{C}_{n,m,\alpha} = \frac{\Delta\beta}{2(1-\alpha)} \sum_{k=1}^m \left(\mathbb{I}\{\beta_{k-1} > \alpha\} \widehat{V}_{n,\beta_{k-1}} + \mathbb{I}\{\beta_k > \alpha\} \widehat{V}_{n,\beta_k} \right). \quad (8)$$

In the next section, we present concentration bounds for the SRM estimator in (7), assuming that the underlying distribution has bounded support, and subsequently, specialize the bound to handle the case of CVaR.

Concentration Bounds

For notational convenience, we shall use V_β and S to denote $V_\beta(X)$ and $S(X)$, for any $\beta \in (0, 1)$.

For all the results presented below, we let $\widehat{S}_{n,m}$ denote the SRM estimate formed using (7). Let F and f denote the distribution and density of X , respectively.

For the sake of analysis, we make the following assumptions on the risk-aversion function and the underlying distribution:

(A1) The function $\varphi(\cdot)$ is differentiable in $\text{int}(\text{supp } \varphi)$, where $\text{int}(\text{supp } \varphi)$ denotes the interior of the support of the function φ . Further, φ satisfies the following conditions: $|\varphi(\beta)| \leq C_1$ and $|\varphi'(\beta)| \leq C_2$, $\forall \beta \in \text{int}(\text{supp } \varphi)$.

(A2) The r.v. X is continuous, with a CDF F that is differentiable on the interior of the support of X , and $|X| \leq B$ a.s. Further, the density f of X satisfies $f(x) \geq 1/\delta_1 > 0$, for $|x| \leq B$.

In (Dowd and Blake 2006), the authors recommend the exponential risk-aversion function, defined by $\varphi(\beta) = \frac{\kappa e^{-\kappa(1-\beta)}}{1-e^{-\kappa}}$, $\beta \in [0, 1]$, for some $\kappa > 0$. This choice for risk-aversion satisfies (A1). Next, in (A2), we assumed that the density f of X is bounded below by $\frac{1}{\delta_1} > 0$. A similar assumption is used for analyzing the classic CVaR estimation scheme, cf. (Kagrecha, Nair, and Jagannathan 2019; Prashanth, Jagannathan, and Kolla 2020; Sun and Hong 2010). Further, one could argue that an assumption like (A2) is necessary to get a meaningful VaR concentration result, since the underlying distribution could be arbitrarily flat — see Remark 3.5 in (Prashanth, Jagannathan, and Kolla 2020). Truncated normal and exponential distributions are examples that satisfy (A2).

The bounds on the error in a trapezoidal approximation of any integral usually assume the derivative of the integrand to be bounded (cf. Corollary 2.2 in (Talvila 2016)). In the context of SRM estimation, this condition translates to a bound on the derivatives of φ and VaR. (A1) ensures $\varphi'(\cdot)$ is bounded, while the condition $f(x) \geq 1/\delta_1 > 0$ in (A2) together with $V'_\beta = \frac{1}{f(V_\beta)}$, $\forall \beta \in (0, 1)$, ensures the derivative of VaR is bounded. Note that (A2) imposes a boundedness assumption on the r.v. X . If the random variable is unbounded, then for every $\epsilon > 0$, there is an x such that $0 < f(x) < \epsilon$. This implies $1/f$ cannot be bounded above uniformly w.r.t x , and hence the derivative of VaR cannot be bounded either.

We now present the main result that establishes an exponential tail bound for the SRM estimator in (7).

Theorem 4.1 (SRM concentration: bounded case). *Assume (A1) and (A2). Choose $K_1 \geq B C_2 + \delta_1 C_1$, where B, δ_1, C_1, C_2 are specified in (A1) and (A2). Fix $\epsilon > 0$. Set the number of sub-intervals $m = \lceil \frac{K_1}{2\epsilon} \rceil$ in (7). Then, we have*

the following bound:

$$\mathbb{P}\left(|S - \widehat{S}_{n,m}| > \epsilon\right) \leq 4 \left\lceil \frac{K_1}{2\epsilon} \right\rceil \exp\left(\frac{-n\epsilon^2}{2\delta_1^2 C_1^2}\right), \quad (9)$$

where δ_1 is as specified in (A2).

Proof. See Section 7. \square

Next, we present a specialization of the tail bound above, to handle the case of CVaR.

Corollary 4.2 (CVaR concentration: bounded case). *Assume (A2). Fix $\alpha \in [0, 1]$, and form the CVaR estimate $\widehat{C}_{n,m,\alpha}$ using (8), with $m = \lceil \frac{K_1}{2\epsilon} \rceil$, where $K_1 \geq \frac{\delta_1}{1-\alpha}$. Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}[|C_\alpha(X) - \widehat{C}_{n,m,\alpha}| > \epsilon] \leq 4 \left\lceil \frac{K_1}{2\epsilon} \right\rceil \exp\left[\frac{-n(1-\alpha)^2\epsilon^2}{2\delta_1^2}\right].$$

Proof. Follows from Theorem 4.1 after observing that the risk-aversion function $\varphi(\cdot)$ defined in (6) is continuous and differentiable in the interior of its support, i.e., on $(\alpha, 1)$. \square

Under stronger conditions on the risk-aversion function and the underlying density, we can obtain an improved concentration result. These conditions are variants of (A1) and (A2), and are formalized below.

(A1') (A1) holds. In addition, $\varphi(\cdot)$ is a twice differentiable function satisfying $|\varphi''(\beta)| \leq C_3$, $\forall \beta \in \text{int}(\text{supp } \varphi)$.

(A2') (A2) holds. In addition, the density $f(\cdot)$ is a differentiable function satisfying $|f'(x)|/f(x)^3 \leq \delta_2$, for $|x| \leq B$.

As before, the exponential risk-aversion function satisfies (A1'). Next, the stronger condition $|f'(x)|/f(x)^3 \leq \delta_2$ used in (A2'), in conjunction with $V''_\beta = -\frac{f'(V_\beta)}{f(V_\beta)^3}$, implies that the second derivative of VaR is bounded. Now, as before, a bounded second derivative implies that the underlying r.v. X is bounded. To see this, the expression for the second derivative of VaR involves a f'/f^3 term, and if the r.v. X is unbounded, then a uniform bound on f'/f^3 would mean that, as $x \rightarrow \infty$, f decays too slowly to integrate to something finite, leading to a contradiction. More precisely, the differential inequality $|f'|/f^3 < K$ can be “solved” to get $f(x) > C/\sqrt{a+bx}$ for large x and suitable constants a, b , and C . However, the expression on the RHS integrates to infinity, and hence, no density f with unbounded support can have f'/f^3 bounded.

Theorem 4.3 (SRM concentration: bounded case (variant)). *Assume (A1') and (A2'). Choose K_2 such that $|B C_3 + 2\delta_1 C_2 + \delta_2 C_1| \leq K_2$, where $B, \delta_1, C_1, C_2, C_3$ are specified in (A1') and (A2'). Fix $\epsilon > 0$. Set the number of sub-intervals $m = \lceil \sqrt{\frac{K_2}{6\epsilon}} \rceil$ in (7). Then, we have the following bound:*

$$\mathbb{P}\left(|S - \widehat{S}_{n,m}| > \epsilon\right) \leq 4 \left\lceil \sqrt{\frac{K_2}{6\epsilon}} \right\rceil \exp\left(\frac{-n\epsilon^2}{2\delta_1^2 C_1^2}\right). \quad (10)$$

Proof. See Section 7. \square

For small values of ϵ , the bound in (10) is better than that in (9). However, the bound in (9) is derived under weaker assumptions on the r.v. X and the risk-aversion function φ , as compared to the bound in (10).

A corollary of the result in Theorem 4.3 can be derived for the case of CVaR, and the reader is referred to (Pandey, Prashanth, and Bhat 2019) for further details.

5 SRM Estimation: Unbounded Case

In this section, we focus on distributions that satisfy an exponential moment bound — an assumption made precise below.

(A3) The r.v. X is continuous, with a differentiable CDF F . Further, there exist $\rho \geq 1$, $\xi > 0$, and $\chi < \infty$ such that $\mathbb{E}(\exp(\xi|X|^\rho)) < \chi$.

Sub-Gaussian and sub-exponential distributions with a differentiable CDF satisfy the assumption above (Vershynin 2018). In particular, a sub-Gaussian r.v. X satisfies (A3) with $\rho = 2$, while a sub-exponential r.v. X satisfies (A3) with $\rho = 1$.

Estimation Scheme

The bounds for the trapezoidal-rule-based approximation of SRM in the section above required the underlying r.v. X to be bounded. Using these bounds together with a truncation-based estimation approach, we control the trapezoidal error approximation for the case of distributions with unbounded support.

Let X_1, \dots, X_n denote i.i.d. samples from the distribution of X . We form a truncated set of samples as follows: for each $i = 1, \dots, n$, define

$$\bar{X}_i = X_i \mathbb{I}\{X_i \leq B_n\}, \text{ with } B_n = \begin{cases} \left[\frac{\rho \log \log n}{\xi(\rho-1)} \right]^{\frac{1}{\rho}}, & \rho > 1, \\ \frac{\log \log n}{\xi}, & \rho = 1. \end{cases} \quad (11)$$

Let $\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\bar{X}_i \leq x]$, $x \in \mathbb{R}$, denote the EDF of the truncated r.v. $\tilde{X} = X \mathbb{I}\{X \leq B_n\}$. Let \tilde{F}_n^{-1} denote the inverse of \tilde{F}_n as defined by (4).

We form an SRM estimate along the lines of (7), except that the samples used are truncated, i.e.,

$$\tilde{S}_{n,m} = \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) \tilde{V}_{n,\beta_{k-1}} + \varphi(\beta_k) \tilde{V}_{n,\beta_k}}{2} \Delta\beta, \quad (12)$$

where $\tilde{V}_{n,\beta} = \tilde{F}_n^{-1}(\beta)$, $\forall \beta \in (0, 1)$.

Concentration Bounds

For the concentration bound presented below, we require the following assumption:

(A4) There exists a positive sequence $\{\psi(n), n \geq 1\}$ such that, for all $n \geq 1$, $\psi(n) = o(\sqrt{n})^1$, and the density f underlying the r.v. X satisfies $f(x) \geq \frac{1}{\psi(n)}$ for all $|x| \leq B_n$, where B_n is defined in (11).

It is apparent that a Gaussian r.v. with variance σ^2 satisfies (A3) with $\rho = 2$, and (A4) with $\psi(n) = \sqrt{2\pi}\sigma (\log n)^{\frac{1}{\sigma^2}}$.

¹ $f(n) = o(g(n))$ if $\frac{f(n)}{g(n)} \rightarrow 0$, as $n \rightarrow \infty$.

Further, an exponential r.v. with parameter λ satisfies (A3) with $\rho = 1$, and (A4) with $\psi(n) = \frac{1}{\lambda} (\log n)^{\frac{\lambda}{\xi}}$.

Theorem 5.1 (SRM concentration: Exponential moment bounded case).

Assume (A1). Suppose the r.v. X with CDF F satisfies (A3) and (A4). Let

$$\tilde{c} = \begin{cases} \frac{\rho\chi C_1}{(e\xi)^{\frac{1}{\rho}}(\rho-1)}, & \rho > 1, \\ \frac{C_1\chi}{\xi} \left(1 + \frac{\chi}{(1-F(B_n))e} \right), & \rho = 1. \end{cases}$$

Fix $n \geq 1$ and $\epsilon > \frac{\tilde{c}}{\log n}$. Let the number m of sub-intervals satisfy $m \geq \frac{(B_n C_2 + \psi(n) C_1)}{(\epsilon - \frac{\tilde{c}}{\log n})}$, where B_n is given by (11). Let $\tilde{S}_{n,m}$ be formed using (12). Then, we have

$$\mathbb{P} \left[\left| S - \tilde{S}_{n,m} \right| > \epsilon \right] \leq \frac{[B_n C_2 + \psi(n) C_1]}{\left(\epsilon - \frac{\tilde{c}}{\log n} \right)} \times \exp \left[- \frac{n \left[\epsilon - \frac{\tilde{c}}{\log n} \right]^2}{2\psi(n)^2 C_1^2} \right].$$

Proof. See Section 7. \square

The bound in the theorem above can be interpreted as follows: For every $\epsilon > 0$, there exists an n_0 such that the tail bound for SRM estimation is applicable for all $n \geq n_0$. It is easy to see that n_0 is the smallest natural number satisfying $n > \exp(\frac{\tilde{c}}{\epsilon})$. We believe that a bound for any estimator on truncation will involve such a minimum number of samples constraint (cf. Proposition 4 in (Bhat and Prashanth 2019)).

For the sub-exponential case, i.e., $\rho = 1$, the tail bound above exhibits exponential decay, while the corresponding bound in (Bhat and Prashanth 2019) shows exponential decay for $\epsilon \leq 1$, and power law decay for $\epsilon > 1$. Further, unlike (Bhat and Prashanth 2019), our bounds make all the constants explicit.

A specialization of the result in 5.1 for the case of CVaR is provided in (Pandey, Prashanth, and Bhat 2019).

6 Simulation Experiments

In this section, we demonstrate the efficacy of our proposed method for SRM estimation (7), which we shall refer to as SRM-Trapz. In our experiments, we use the risk aversion function $\varphi(\beta) = \frac{5e^{-5(1-\beta)}}{1-e^{-5}}$, $\beta \in [0, 1]$. In the following sub-section, we consider a synthetic experimental setting to compare the accuracy of SRM estimators. Subsequently, we use SRM-Trapz as a subroutine in a vehicular traffic routing application (see section 6).

Synthetic Setup

Figure 1 presents the estimation error as a function of the sample size for SRM-Trapz. The algorithm is run with two different sub-divisions. The samples are generated using a Gaussian distribution with mean 0.5 and variance 25. We observe that SRM-Trapz with 500 subdivisions performs on

par with SRM-Trapz with 150 subdivisions for every sample size. Further, as expected, increasing sample size leads to lower estimation error, while also increasing the confidence (demonstrated by the shrinkage in standard error).

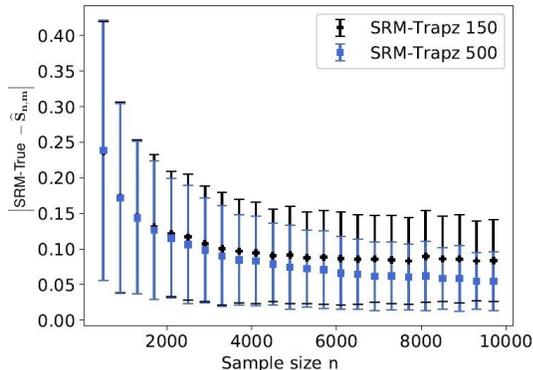


Figure 1: Error in SRM estimation ($|\text{SRM-True} - \text{Empirical SRM}|$) for different sample sizes. SRM-True is approximated using (7) with a large number of samples. Empirical SRM is calculated by two methods, (i) SRM-Trapz method with $m = 150$ subdivisions (SRM-Trapz 150), and (ii) SRM-Trapz method with $m = 500$ subdivisions (SRM-Trapz 500). In both methods, SRM is estimated using (7). The underlying distribution considered for this simulation is $X \sim \mathcal{N}(0.5, 5^2)$. The bars in the plot shows standard error averaged over 10^3 trials.

Distribution	SRM-True	SRM 1000
Exp(0.2)	10.99±0.01	11.02±1.21
$\mathcal{N}(0, 10^2)$	107.36±0.01	107.80±1.32
Exp(0.01)	221.30±0.01	221.39±2.47
$\mathcal{U}(-10^3, 10^3)$	612.47±0.02	612.65±4.91

Table 1: The results for SRM estimation, on four distributions, using two methods. Distributions are (a) Exponential distribution with mean $1/0.2$ (Exp(0.2)), (b) Normal distribution with mean zero and variance 10^2 ($\mathcal{N}(0, 10^2)$), (c) Exponential distribution with mean $1/0.01$ (Exp(0.01)), (d) Uniform distribution with range -10^3 to 10^3 ($\mathcal{U}(-10^3, 10^3)$). Methods are (i) Calculation of SRM (SRM-True) using (7) with $m = 10^5$ subdivisions, (ii) SRM-Trapz method (SRM-Trapz 1000) using (7) with $m = 10^3$ subdivisions. In method (i), 10^{12} and in method (ii), 10^4 i.i.d. samples are used for estimating SRM on each distribution, and the standard error is averaged over 10^6 and 10^3 iterations respectively.

Table 1 presents the results obtained by SRM-Trapz with 1000 subdivisions, for four different input distributions, three of which have unbounded support. SRM-Trapz, which is the un-truncated estimator, is found to exhibit low estimation error even for distributions with unbounded support. In Table 1, SRM-True is used as a proxy for the true SRM value in the experiments, and is set using (7) with a large number

(approximately, 10^{12}) of samples. Given the lack of a closed form expression for SRM, the estimation scheme that we propose is a viable alternative to form a proxy SRM-True for the true SRM value.

Vehicular Traffic Routing

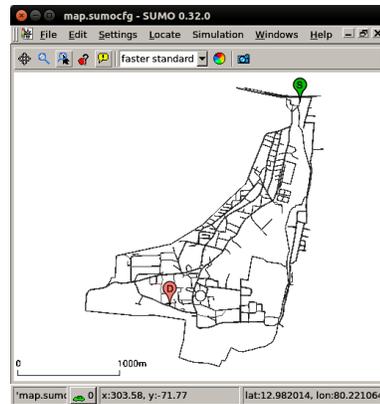


Figure 2: Area of an urban city map, used for SUMO network.

In the vehicular routing application, the traditional objective is to find a route with the lowest expected delay. However, such an objective ignores risk factors. An alternative is to consider the weighted-sum delay of each route, and we use SRM to quantify this objective. Thus, given a set of routes, the aim is to find the route (by adaptive sampling) with the lowest SRM of the delay. Simulation of Urban MObility (SUMO) (Behrisch et al. 2011) is an open source, highly portable, microscopic road traffic simulation package designed to handle large road networks. Traffic Control Interface (TraCI) (Wenger et al. 2008) is a library, providing extensive commands to control the behavior of the simulation online, including vehicle state, road configuration, and traffic lights. We implement our routing algorithm using SUMO and TRACI.

For the experiments, we use the street map of the area around IIT Madras, Chennai, India (see Figure 2) obtained from OpenStreetMap (OSM) (Haklay and Weber 2008), and then used Netconvert tool to load the map in SUMO. The network has 426 junctions and a total edge length of 123 km. We ran SUMO on this network for 30,000 time-steps, in which 7000 cars, 500 buses, 2000 bikes, 1000 cycles, and 1000 pedestrians were added at different time-steps and in different lanes uniformly. We choose $K = 5$ routes between two fixed points, marked as S and D in Figure 2. On these selected routes, we added $n = 1000$ cars and tracked them. In Table 2, $\hat{X}_{n,i}$ is the estimated average delay of the i th route, and $\hat{S}_{n,m,i}$ is the SRM estimate for the i th route, $i = 1, \dots, K$, using (7), and with n samples. The estimates are averages over 1000 independent trials. We set the number of subdivisions $m = 100$.

From Table 2 it is apparent that ROUTE₄ has the lowest expected delay, and ROUTE₂ has the lowest SRM. We consider a best-arm identification (BAI) bandit framework (Audibert, Bubeck, and Munos 2010), where an algorithm is given a fixed budget. Here, the budget refers to the total number of

	ROUTE ₁	ROUTE ₂	ROUTE ₃	ROUTE ₄	ROUTE ₅
$\hat{X}_{n,i}$	283.81	287.15	306.80	266.85	325.86
$\hat{S}_{n,m,i}$	431.28	361.81	455.83	378.68	390.95

Table 2: Results for the estimated average delay ($\hat{X}_{n,i}$) and estimated SRM ($\hat{S}_{n,m,i}$), for i th route, where $i = 1, \dots, K$.

samples across routes. After the sampling budget, the algorithm is expected to recommend a route, and is judged by the probability that the recommended route is correct (i.e. the best route). We ran successive rejects (SR) (Audibert, Bubeck, and Munos 2010), which is a popular BAI algorithm, except that SR is modified to find the route with lowest SRM. Note that the regular SR algorithm finds the route with the lowest expected delay. The setting of SUMO is as noted above. We set the budget $n = 1000$, number of routes $K = 5$, and $m = 100$ subdivisions for SRM-Trapz. We observed that SR algorithm picks ROUTE₂ with probability 0.91.

7 Convergence Proofs

Proof of Theorem 4.1

We establish two results concerning the error of a trapezoidal-rule-based approximation, and a concentration bound for the VaR estimate in (4). These results will be used in the proof of Theorem 4.1, which is provided subsequently.

Lemma 7.1. *Under conditions of Theorem 4.1, we have*

$$\left| \int_0^1 \varphi(\beta) V_\beta d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta \right| \leq \frac{K_1}{4m}. \quad (13)$$

Proof. Let $\mathcal{L}^\infty([0, 1])$ denote the space of essentially bounded functions on $[0, 1]$. For any $f \in \mathcal{L}^\infty([0, 1])$, let $\|f\|_\infty = \inf\{\tau \mid |f(x)| \leq \tau \text{ for almost every } x \in [0, 1]\}$ denote the associated norm.

Then, for $k = 1, \dots, m$, we have

$$\begin{aligned} \mathcal{E}_k &\triangleq \left| \int_{\beta_k}^{\beta_{k+1}} \varphi(\beta) V_\beta d\beta - \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta \right| \\ &\leq \frac{1}{4m^2} \left\| (\varphi(\cdot) V_{(\cdot)})' \right\|_\infty \leq \frac{K_1}{4m^2}, \end{aligned} \quad (14)$$

where the first inequality in (14) follows from Corollary 2.2 in (Talvila 2016). The latter result provides a bound on the error in the trapezoidal approximation of a Lebesgue integral. For the second inequality in (14), we have used (A1), (A2), and the choice of K_1 to infer

$$\left\| (\varphi(\cdot) V_{(\cdot)})' \right\|_\infty \leq B C_2 + \delta_1 C_1 \leq K_1.$$

The main claim can be inferred as follows:

$$\left| \int_0^1 \varphi(\beta) V_\beta d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta \right| \leq \sum_{k=1}^m \mathcal{E}_k \leq m \times \frac{K_1}{4m^2} = \frac{K_1}{4m}.$$

□

Lemma 7.2 (VaR concentration). *Assume (A2). Then, for any $\beta \in (0, 1)$ and $\epsilon > 0$, we have*

$$\mathbb{P} \left[\left| V_\beta - \hat{V}_{n,\beta} \right| \geq \epsilon \right] \leq 2 \exp \left(-\frac{2n\epsilon^2}{\delta_1^2} \right), \quad (15)$$

where δ_1 is specified in (A2).

Proof. Following the proof of Proposition 2 in (Kolla et al. 2019), we obtain

$$\mathbb{P} \left(\left| \hat{V}_{n,\beta} - V_\beta \right| \geq \epsilon \right) \leq 2 \exp \left(-2n\zeta_\epsilon^2 \right), \quad (16)$$

where $\zeta_\epsilon = \min\{F(V_\alpha + \epsilon) - F(V_\alpha), F(V_\alpha) - F(V_\alpha - \epsilon)\}$. Using (A2), we have $F(V_\alpha + \epsilon) - F(V_\alpha) = f(\bar{V})\epsilon \geq \frac{\epsilon}{\delta_1}$, for some $\bar{V} \in (V_\alpha, \min\{V_\alpha + \epsilon, B\})$. Along similar lines, we have $F(V_\alpha) - F(V_\alpha - \epsilon) \geq \frac{\epsilon}{\delta_1}$, and we obtain $\zeta_\epsilon \geq \frac{\epsilon}{\delta_1}$. The main claim follows by using the inequality above in (16). □

Proof of Theorem 4.1. Notice that

$$\begin{aligned} &\mathbb{P} \left[\left| S - \hat{S}_{n,m} \right| > \epsilon \right] \\ &\leq \mathbb{P} \left[\left| \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) \hat{V}_{n,\beta_{k-1}} + \varphi(\beta_k) \hat{V}_{n,\beta_k}}{2} \Delta\beta \right| > \frac{\epsilon}{2} \right]. \end{aligned}$$

The inequality above follows by using Lemma 7.1 to infer that for $m \geq \frac{K_1}{2\epsilon}$, we have $\left| \int_0^1 \varphi(\beta) V_\beta d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta \right| < \frac{\epsilon}{2}$. Now, we have

$$\begin{aligned} &\mathbb{P} \left[\left| S - \hat{S}_{n,m} \right| > \epsilon \right] \\ &\leq \mathbb{P} \left[\left| \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) V_{\beta_{k-1}} + \varphi(\beta_k) V_{\beta_k}}{2} \Delta\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1}) \hat{V}_{n,\beta_{k-1}} + \varphi(\beta_k) \hat{V}_{n,\beta_k}}{2} \Delta\beta \right| > \frac{\epsilon}{2} \right] \\ &\leq \mathbb{P} \left[\left| \varphi(\beta_0) V_{\beta_0} - \varphi(\beta_0) \hat{V}_{n,\beta_0} \right| > \frac{\epsilon}{2m \Delta\beta} \right] \\ &\quad + 2\mathbb{P} \left[\left| \varphi(\beta_1) V_{\beta_1} - \varphi(\beta_1) \hat{V}_{n,\beta_1} \right| > \frac{\epsilon}{2m \Delta\beta} \right] + \dots + \\ &\quad 2\mathbb{P} \left[\left| \varphi(\beta_{m-1}) V_{\beta_{m-1}} - \varphi(\beta_{m-1}) \hat{V}_{n,\beta_{m-1}} \right| > \frac{\epsilon}{2m \Delta\beta} \right] \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left[\left| \varphi(\beta_m) V_{\beta_m} - \varphi(\beta_m) \widehat{V}_{n, \beta_m} \right| > \frac{\epsilon}{2m \Delta \beta} \right] \\
& \leq 2 \exp \left(-\frac{2n}{\delta_1^2} \left(\frac{\epsilon}{2m \varphi(\beta_0) \Delta \beta} \right)^2 \right) \\
& \quad + 4 \exp \left(-\frac{2n}{\delta_1^2} \left(\frac{\epsilon}{2m \varphi(\beta_1) \Delta \beta} \right)^2 \right) \\
& \quad + \cdots + 4 \exp \left(-\frac{2n}{\delta_1^2} \left(\frac{\epsilon}{2m \varphi(\beta_{m-1}) \Delta \beta} \right)^2 \right) \\
& \quad + 2 \exp \left(-\frac{2n}{\delta_1^2} \left(\frac{\epsilon}{2m \varphi(\beta_m) \Delta \beta} \right)^2 \right),
\end{aligned}$$

where we applied Lemma 7.2 to obtain the final inequality. Thus,

$$\begin{aligned}
\mathbb{P} \left[\left| S - \widehat{S}_{n, m} \right| > \epsilon \right] & \leq 4m \exp \left(-\frac{2n}{\delta_1^2} \left(\frac{\epsilon}{2m C_1 \Delta \beta} \right)^2 \right) \\
& = 4m \exp \left(-\frac{n \epsilon^2}{2 \delta_1^2 C_1^2} \right) = \left\lceil \frac{2K_1}{\epsilon} \right\rceil \exp \left(-\frac{n \epsilon^2}{2 \delta_1^2 C_1^2} \right). \tag{17}
\end{aligned}$$

The claim follows. \square

Proof of Theorem 4.3. The proof follows in a similar manner as that of Theorem 4.1, using a variation of the result in Lemma 7.1. The reader is referred to Section 6 of (Pandey, Prashanth, and Bhat 2019) for the details. \square

Proof of Theorem 5.1

Proof. Notice that

$$\mathbb{P} \left[S - \widetilde{S}_{n, m} > \epsilon \right] = \mathbb{P} \left[I_1 + I_2 > \epsilon \right], \text{ where}$$

$$I_1 = \int_0^\eta \varphi(\beta) V_\beta d\beta - \sum_{k=1}^m \frac{\varphi[\beta_{k-1}] \widetilde{V}_{n, \beta_{k-1}} + \varphi[\beta_k] \widetilde{V}_{n, \beta_k}}{2} \Delta \beta,$$

$$\text{and } I_2 = \int_\eta^1 \varphi(\beta) V_\beta d\beta.$$

In the above, $\eta = F(B_n)$, with B_n as defined in the theorem statement. We bound I_2 as follows:

$$\begin{aligned}
1 - \beta & = \mathbb{P}(X > V_\beta) = \mathbb{P} \left(\exp(\xi X^\rho) > \exp(\xi V_\beta^\rho) \right) \\
& \leq \chi \exp \left(-\xi V_\beta^\rho \right), \tag{18}
\end{aligned}$$

$$\text{or equivalently, } V_\beta \leq \left(\frac{1}{\xi} \log \left(\frac{\chi}{1 - \beta} \right) \right)^{\frac{1}{\rho}}. \tag{19}$$

We now handle the case of $\rho > 1$ and $\rho = 1$ separately.

Case $\rho > 1$:

Using $\log x \leq \frac{x}{e} \forall x > 0$, we obtain $V_\beta \leq \left(\frac{\chi}{e \xi (1 - \beta)} \right)^{\frac{1}{\rho}}$, leading to

$$\int_\eta^1 V_\beta d\beta \leq \left(\frac{\chi}{e \xi} \right)^{\frac{1}{\rho}} \int_\eta^1 \frac{d\beta}{(1 - \beta)^{\frac{1}{\rho}}} = \left(\frac{\chi}{e \xi} \right)^{\frac{1}{\rho}} \frac{\rho(1 - \eta)^{1 - \frac{1}{\rho}}}{\rho - 1}$$

$$\begin{aligned}
& \leq \left(\frac{1}{e \xi} \right)^{\frac{1}{\rho}} \frac{\rho}{\rho - 1} \chi \exp \left(-\frac{\xi V_\eta^\rho (\rho - 1)}{\rho} \right) \tag{using (18)} \\
& = \left(\frac{1}{e \xi} \right)^{\frac{1}{\rho}} \frac{\rho}{\rho - 1} \chi \exp \left(-\frac{\xi B_n^\rho (\rho - 1)}{\rho} \right) \tag{since } V_\eta = B_n
\end{aligned}$$

Hence,

$$I_2 \leq C_1 \int_\eta^1 V_\beta d\beta \leq \frac{\rho \chi C_1}{(e \xi)^{\frac{1}{\rho}} (\rho - 1) \log n}, \tag{20}$$

where the final inequality uses the definition of B_n in (11).

Case $\rho = 1$:

Using (19), we have

$$\begin{aligned}
\int_\eta^1 V_\beta d\beta & \leq \frac{1}{\xi} \int_\eta^1 \log \left(\frac{\chi}{1 - \beta} \right) d\beta \\
& \leq \frac{(1 - \eta)}{\xi} \left[1 + \frac{\chi}{(1 - \eta)e} \right] \leq \frac{\chi \exp(-\xi V_\eta)}{\xi} \left[1 + \frac{\chi}{(1 - \eta)e} \right] \\
& = \frac{\chi \exp(-\xi B_n)}{\xi} \left(1 + \frac{\chi}{(1 - \eta)e} \right). \tag{since } V_\eta = B_n
\end{aligned}$$

Using (A1) and $B_n = \frac{\log \log(n)}{\xi}$, we obtain

$$I_2 \leq C_1 \int_\eta^1 V_\beta d\beta \leq \frac{C_1 \chi}{\xi \log n} \left(1 + \frac{\chi}{(1 - \eta)e} \right) \tag{21}$$

Applying the bound in the Theorem 4.1 to the truncated r.v. $Z = X \mathbb{I} \{X \leq B_n\}$, and using (A4), we obtain

$$\mathbb{P} [I_1 > \epsilon] \leq \frac{K_1}{\epsilon} \exp \left(-\frac{n \epsilon^2}{2 \psi(n)^2 C_1^2} \right). \tag{22}$$

With \tilde{c} as in the theorem statement, we obtain

$$\begin{aligned}
& \mathbb{P} [I_1 + I_2 > \epsilon] \\
& \leq \frac{K_1}{\left(\epsilon - \frac{\tilde{c}}{\log n} \right)} \exp \left(-\frac{n \left(\epsilon - \frac{\tilde{c}}{\log n} \right)^2}{2 \psi(n)^2 C_1^2} \right) \\
& \quad \text{(using (20), (21) and (22))} \\
& = \frac{(B_n C_2 + \psi(n) C_1)}{\left(\epsilon - \frac{\tilde{c}}{\log n} \right)} \exp \left(-\frac{n \left(\epsilon - \frac{\tilde{c}}{\log n} \right)^2}{2 \psi(n)^2 C_1^2} \right).
\end{aligned}$$

By using a parallel argument, a concentration result for bounding the lower semi-deviations can be derived. We omit the details. \square

8 Conclusions

We proposed a novel SRM estimation scheme that is based on numerical integration, and derived concentration bounds for our SRM estimator for the case of distributions with bounded support, as well as distributions satisfying an exponential moment bound. As future work, it would be interesting to derive a lower bound for SRM estimation, and close the gap (if any) w.r.t. the upper bound that we have derived.

Acknowledgments

This research work was supported in part by a Department of Science and Technology (DST) grant under the ECRA program.

References

- Acerbi, C. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance* 26(7): 1505–1518.
- Acerbi, C.; and Tasche, D. 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26(7): 1487–1503.
- Artzner, P.; Delbaen, F.; Eber, J. M.; and Heath, D. 1999. Coherent measures of risk. *Mathematical finance* 9(3): 203–228.
- Audibert, J. Y.; Bubeck, S.; and Munos, R. 2010. Best arm identification in multi-armed bandits. In *Conference on Learning Theory*, 41–53.
- Behrisch, M.; Bieker, L.; Erdmann, J.; and Krajzewicz, D. 2011. SUMO—Simulation of Urban MObility. In *The Third International Conference on Advances in System Simulation (SIMUL 2011)*, volume 42.
- Bhat, S. P.; and Prashanth, L. A. 2019. Concentration of risk measures: A Wasserstein distance approach. In *Advances in Neural Information Processing Systems*, 11739–11748.
- Brown, D. B. 2007. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters* 35(6): 722–730.
- Dowd, K.; and Blake, D. 2006. After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance* 73(2): 193–229.
- Galichet, N.; Sebag, M.; and Teytaud, O. 2013. Exploration vs exploitation vs safety: Risk-aware multi-armed bandits. In *Asian Conference on Machine Learning*, 245–260.
- Haklay, M.; and Weber, P. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7(4): 12–18.
- Kagrecha, A.; Nair, J.; and Jagannathan, K. 2019. Distribution oblivious, risk-aware algorithms for multi-armed bandits with unbounded rewards. In *Advances in Neural Information Processing Systems*, 11269–11278.
- Khim, J.; Leqi, L.; Prasad, A.; and Ravikumar, P. 2020. Uniform convergence of rank-weighted learning. In *International Conference on Machine Learning*, 5254–5263. PMLR.
- Kolla, R. K.; Prashanth, L. A.; Bhat, S. P.; and Jagannathan, K. P. 2019. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters* 47(1): 16–20.
- Krokhmal, P.; Palmquist, J.; and Uryasev, S. 2002. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4: 43–68.
- Mulvey, J. M.; and Erkan, H. G. 2006. Applying CVaR for decentralized risk management of financial companies. *Journal of Banking & Finance* 30(2): 627–644.
- Pandey, A. K.; Prashanth, L. A.; and Bhat, S. P. 2019. Estimation of Spectral Risk Measures. *CoRR* abs/1912.10398.
- Prashanth, L. A.; Jagannathan, K.; and Kolla, R. K. 2020. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions. In *International Conference on Machine Learning*, volume 119, 5577–5586. PMLR.
- Serfling, R. J. 2009. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons.
- Sun, L.; and Hong, L. J. 2010. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters* 38(4): 246–251.
- Talvila, E. 2016. Higher order corrected trapezoidal rules in Lebesgue and Alexiewicz spaces. *arXiv preprint arXiv:1604.08643*.
- Vershynin, R. 2018. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.
- Wainwright, M. J. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.
- Wang, Y.; and Gao, F. 2010. Deviation inequalities for an estimator of the conditional value-at-risk. *Operations Research Letters* 38(3): 236–239.
- Wegener, A.; et al. 2008. TraCI: an interface for coupling road traffic and network simulators. In *Proceedings of the 11th communications and networking simulation symposium*, 155–163. ACM.